Valuing Financial Data

by Farboodi, Singal, Veldkamp, and Venkateswaran

Discussion by Cecilia Parlatore

NYU Stern

July 15, 2022

Overview

Broad question: What is the value of information?

Overview

Broad question: What is the value of information?

- Depends on use of information, preferences, priors/existing information, who has access to information, etc.
- Hard question to answer! Even harder to measure.

Overview

Broad question: What is the value of information?

- Depends on use of information, preferences, priors/existing information, who has access to information, etc.
- Hard question to answer! Even harder to measure.

This paper How *should* an investor value financial data?

- Private value for an individual investor - bypass aggregation, asset pricing
- Can restrict use of data: Investors trade a specific set of assets

This paper

What does this paper do?

- Construct a theoretically-based measure of the value on information for an investor based on
  1. moments of the unconditional return distribution
  2. conditional return volatility
  3. risk aversion and price impact
- Main application: estimate value of IBES forecast data
- More applications in the paper + many potential others

This paper

What does this paper do?

- Construct a theoretically-based measure of the value on information for an investor based on
    1. moments of the unconditional return distribution
    2. conditional return volatility
    3. risk aversion and price impact
- Main application: estimate value of IBES forecast data
- More applications in the paper + many potential others

Contribution: methodology to value data

1. Approximate the economy with a standard NREE model
2. Compute the value of data (in the approximated model)
3. Use the structure of the model to estimate the value of data

Measuring the value of data

- Key step in methodology: second order approximation of utility

$$u_i(c) \approx u_i\left(\mathbb{E}^i\left[\tilde{c_i}(\mathcal{I})\right]\right) + \frac{u_i'\left(\mathbb{E}^i\left[c_i(\mathcal{I})\right]\right)}{\rho_i} \underbrace{\left[\rho_i\left(c - \mathbb{E}^i\left[c_i(\mathcal{I})\right]\right) + \frac{\rho_i^2}{2}\left(c - \mathbb{E}^i\left[c_i(\mathcal{I})\right]\right)^2\right]}_{\equiv U_i(c)}$$

  where $\rho_i = \rho_i(w_0, \mathcal{I})$ is the absolute risk aversion evaluated at $\mathbb{E}^i\left[c_i(\mathcal{I})\right]$

- Advantage: working with U_i gives closed forms for indirect utilities as functions of things we can (mostly) measure from the data

My discussion

1. Interpretation of estimates: timing of approximation
   - Value of data in approximated economy $\neq$ Approximated value of data
2. From theory to measurement: highlight assumptions to compute the value of data
3. Measurement: Can we exploit the structure more?

Timing of approximation

- We want to compute the **dollar value** $\Delta^\star$ of data $\mathcal{Z}$ for an investor i
- Investor i is indifferent between paying $\Delta^\star$ dollars to get $\mathcal{Z}$ and not doing so

$$\mathbb{E}\left[u_i\left(\mathcal{I},\mathcal{Z};w_0-\Delta^\star\right)\right]=\mathbb{E}\left[u_i\left(\mathcal{I};w_0\right)\right]$$

- Willingness to pay for data $\mathcal{Z}$

$$\Delta^\star \approx \frac{\mathbb{E}\left[u_i\left(\mathcal{I},\mathcal{Z};w_0\right)\right]-\mathbb{E}\left[u_i\left(\mathcal{I};w_0\right)\right]}{r\mathbb{E}\left[u_i'\left(c\left(\mathcal{I},\mathcal{Z}\right)\right)\right]}$$

Timing of approximation

- We want to compute the **dollar value** $\Delta^\star$ of data $\mathcal{Z}$ for an investor i
- Investor i is indifferent between paying $\Delta^\star$ dollars to get $\mathcal{Z}$ and not doing so

$$\mathbb{E}\left[u_i\left(\mathcal{I},\mathcal{Z};w_0-\Delta^\star\right)\right]=\mathbb{E}\left[u_i\left(\mathcal{I};w_0\right)\right]$$

- Willingness to pay for data $\mathcal{Z}$

$$\Delta^\star\approx\frac{\mathbb{E}\left[u_i\left(\mathcal{I},\mathcal{Z};w_0\right)\right]-\mathbb{E}\left[u_i\left(\mathcal{I};w_0\right)\right]}{r\mathbb{E}\left[u_i'\left(c\left(\mathcal{I},\mathcal{Z}\right)\right)\right]}$$

- Proposed measure based on approximated utility

$$\Delta=\frac{\mathbb{E}\left[U_i\left(\mathcal{I},\mathcal{Z}\right)\right]-\mathbb{E}\left[U_i\left(\mathcal{I}\right)\right]}{r\mathbb{E}\left[U_i'\left(c\left(\mathcal{I}\right)\right)\right]}$$

Timing of approximation

- We want to compute the **dollar value** $\Delta^\star$ of data $\mathcal{Z}$ for an investor i
- Investor i is indifferent between paying $\Delta^\star$ dollars to get $\mathcal{Z}$ and not doing so

$$\mathbb{E}\left[u_i\left(\mathcal{I},\mathcal{Z};w_0-\Delta^\star\right)\right]=\mathbb{E}\left[u_i\left(\mathcal{I};w_0\right)\right]$$

- Willingness to pay for data $\mathcal{Z}$

$$\Delta^\star\approx\frac{\mathbb{E}\left[u_i\left(\mathcal{I},\mathcal{Z};w_0\right)\right]-\mathbb{E}\left[u_i\left(\mathcal{I};w_0\right)\right]}{r\mathbb{E}\left[u_i'\left(c\left(\mathcal{I},\mathcal{Z}\right)\right)\right]}$$

- Proposed measure based on approximated utility

$$\Delta=\frac{\mathbb{E}\left[U_i\left(\mathcal{I},\mathcal{Z}\right)\right]-\mathbb{E}\left[U_i\left(\mathcal{I}\right)\right]}{r\mathbb{E}\left[U_i'\left(c\left(\mathcal{I}\right)\right)\right]}$$

- How good is Δ as an approximation of $\Delta^\star$?

How good is Δ as an approximation of $\Delta^\star$?

- Value of $\mathcal{Z}$ measured in dollars

$$\Delta^\star = \frac{\mathbb{E}^i\left[u_i'\left(c_i\left(\mathcal{I}\right)\right)\right]}{\mathbb{E}\left[u_i'\left(c\left(\mathcal{I},\mathcal{Z}\right)\right)\right]} \frac{u_i'\left(\mathbb{E}^i\left[\tilde{c_i}\left(\mathcal{I}\right)\right]\right)}{\mathbb{E}^i\left[u_i'\left(c_i\left(\mathcal{I}\right)\right)\right]} \frac{\mathbb{E}^i\left[U_i\left(\mathcal{I},\mathcal{Z}\right)\right]-\mathbb{E}^i\left[U_i\left(\mathcal{I}\right)\right]}{r\mathbb{E}\left[U_i'\left(c\left(\mathcal{I}\right)\right)\right]}$$

How good is Δ as an approximation of $\Delta^\star$?

- Value of $\mathcal{Z}$ measured in dollars

$$\Delta^\star = \frac{\mathbb{E}^i\left[u_i'\left(c_i\left(\mathcal{I}\right)\right)\right]}{\mathbb{E}\left[u_i'\left(c\left(\mathcal{I},\mathcal{Z}\right)\right)\right]}\frac{u_i'\left(\mathbb{E}^i\left[\tilde{c_i}\left(\mathcal{I}\right)\right]\right)}{\mathbb{E}^i\left[u_i'\left(c_i\left(\mathcal{I}\right)\right)\right]}\Delta$$

How good is Δ as an approximation of $\Delta^\star$?

- Value of $\mathcal{Z}$ measured in dollars

$$\Delta^\star = \frac{\mathbb{E}^i\left[u_i'\left(c_i\left(\mathcal{I}\right)\right)\right]}{\mathbb{E}\left[u_i'\left(c\left(\mathcal{I},\mathcal{Z}\right)\right)\right]}\frac{u_i'\left(\mathbb{E}^i\left[\tilde{c_i}\left(\mathcal{I}\right)\right]\right)}{\mathbb{E}^i\left[u_i'\left(c_i\left(\mathcal{I}\right)\right)\right]}\Delta$$

- Goodness of fit depends on
  - Approximation of marginal utility of wealth likely > 1
  - Approximation of utility function > 1 iff $u''' < 0$

How good is Δ as an approximation of $\Delta^\star$?

- Value of $\mathcal{Z}$ measured in dollars

$$\Delta^\star = \frac{\mathbb{E}^i\left[u_i'\left(c_i\left(\mathcal{I}\right)\right)\right]}{\mathbb{E}\left[u_i'\left(c\left(\mathcal{I},\mathcal{Z}\right)\right)\right]} \frac{u_i'\left(\mathbb{E}^i\left[\tilde{c_i}\left(\mathcal{I}\right)\right]\right)}{\mathbb{E}^i\left[u_i'\left(c_i\left(\mathcal{I}\right)\right)\right]}\Delta$$

- Goodness of fit depends on
  - Approximation of marginal utility of wealth likely > 1
  - Approximation of utility function > 1 iff $u''' < 0$
- Interpret the estimate in the paper as a lower bound
- Need more structure to go on!

Lower bound on value of data

$$\Delta = \frac{\mathbb{E}^{i}\left[U_{i}\left(\mathcal{I},\mathcal{Z}\right)\right] - \mathbb{E}^{i}\left[U_{i}\left(\mathcal{I}\right)\right]}{r\mathbb{E}\left[U_{i}'\left(c\left(\mathcal{I}\right)\right)\right]}$$

- Exploit full structure of NREE model to compute indirect utilities
  - Linear structure of equilibrium $\Rightarrow$ all random variables are normal!
  - Can explicitly characterize all distributions
  - Can get a lot of mileage from the structure to go from theory to measurement!

From theory to measurement

$$\Delta \approx \frac{1}{2\rho r}\left(\underbrace{\mathbb{E}^i[\Pi]\left[(\mathcal{V}_i(\mathcal{I},\mathcal{Z}))^{-1}-(\mathcal{V}_i(\mathcal{I}))^{-1}\right]\mathbb{E}^i[\Pi]'}_{\text{change in risk adjusted profit}}+\underbrace{\mathbb{V}^i[\Pi]\left[(\mathcal{V}_i(\mathcal{I},\mathcal{Z}))^{-1}-(\mathcal{V}_i(\mathcal{I}))^{-1}\right]}_{\text{variance reduction}}\right)$$

where Π is a vector of excess returns

- Need numbers for
  - investors priors $\mathbb{E}^i[\Pi]$, $\mathbb{V}^i[\Pi]$
  - absolute risk aversion ρ
  - $\left[(\mathcal{V}_i(\mathcal{I},\mathcal{Z}))^{-1}-(\mathcal{V}_i(\mathcal{I}))^{-1}\right]$

From theory to measurement

$$\Delta \approx \frac{1}{2\rho r}\left(\underbrace{\mathbb{E}^i\left[\Pi\right]\left[\left(\mathcal{V}_i\left(\mathcal{I},\mathcal{Z}\right)\right)^{-1}-\left(\mathcal{V}_i\left(\mathcal{I}\right)\right)^{-1}\right]\mathbb{E}^i\left[\Pi\right]'}_{\text{change in risk adjusted profit}}+\underbrace{\mathbb{V}^i\left[\Pi\right]\left[\left(\mathcal{V}_i\left(\mathcal{I},\mathcal{Z}\right)\right)^{-1}-\left(\mathcal{V}_i\left(\mathcal{I}\right)\right)^{-1}\right]}_{\text{variance reduction}}\right)$$

where Π is a vector of excess returns

- Need numbers for
  - investors priors $\mathbb{E}^i\left[\Pi\right]$, $\mathbb{V}^i\left[\Pi\right]$ $\Rightarrow$beliefs = econometrician's beliefs: get $\mathbb{E}\left[\Pi\right]$, $\mathbb{V}\left[\Pi\right]$ from data
  - absolute risk aversion ρ
  - $\left[\left(\mathcal{V}_i\left(\mathcal{I},\mathcal{Z}\right)\right)^{-1}-\left(\mathcal{V}_i\left(\mathcal{I}\right)\right)^{-1}\right]$

From theory to measurement

$$\Delta \approx \frac{1}{2\rho r}\left(\underbrace{\mathbb{E}^i[\Pi]\left[(\mathcal{V}_i(\mathcal{I},\mathcal{Z}))^{-1}-(\mathcal{V}_i(\mathcal{I}))^{-1}\right]\mathbb{E}^i[\Pi]'}_{\text{change in risk adjusted profit}}+\underbrace{\mathbb{V}^i[\Pi]\left[(\mathcal{V}_i(\mathcal{I},\mathcal{Z}))^{-1}-(\mathcal{V}_i(\mathcal{I}))^{-1}\right]}_{\text{variance reduction}}\right)$$

where Π is a vector of excess returns

- Need numbers for
  - investors priors $\mathbb{E}^i[\Pi]$, $\mathbb{V}^i[\Pi]$ $\Rightarrow$beliefs = econometrician's beliefs: get $\mathbb{E}[\Pi]$, $\mathbb{V}[\Pi]$ from data
  - absolute risk aversion ρ $\Rightarrow$ calibrate it to have CRRA$= 2$
  - $\left[(\mathcal{V}_i(\mathcal{I},\mathcal{Z}))^{-1}-(\mathcal{V}_i(\mathcal{I}))^{-1}\right]$: two cases for $\mathcal{V}_i$ depending on investor's market power

Competitive vs Strategic investor

- Competitive investor

$$\left(\mathcal{V}_i\left(\mathcal{I},\mathcal{Z}\right)\right)^{-1} - \left(\mathcal{V}_i\left(\mathcal{I}\right)\right)^{-1} = \underbrace{\left(\mathbb{V}\left[\Pi|\mathcal{I},\mathcal{Z}\right]\right)^{-1} - \left(\mathbb{V}\left[\Pi|\mathcal{I}\right]\right)^{-1}}_{\text{precision of information in } \mathcal{Z}}$$

  - Exploiting Bayesian updating and linearity can estimate using OLS regressions

Competitive vs Strategic investor

- Competitive investor

$$(\mathcal{V}_i(\mathcal{I},\mathcal{Z}))^{-1} - (\mathcal{V}_i(\mathcal{I}))^{-1} = \underbrace{(\mathbb{V}[\Pi|\mathcal{I},\mathcal{Z}])^{-1} - (\mathbb{V}[\Pi|\mathcal{I}])^{-1}}_{\text{precision of information in } \mathcal{Z}}$$

  - Exploiting Bayesian updating and linearity can estimate using OLS regressions

- Investor with market power

$$(\mathcal{V}_i(\mathcal{I},\mathcal{Z}))^{-1} - (\mathcal{V}_i(\mathcal{I}))^{-1} = F\left(\mathbb{V}^i[\Pi|\mathcal{I}], \mathbb{V}^i[\Pi|\mathcal{I},\mathcal{Z}], \frac{1}{\rho}\left.\frac{\partial \ln[p]}{\partial q}\right|_{\mathcal{I}}, \frac{1}{\rho}\left.\frac{\partial \ln[p]}{\partial q}\right|_{\mathcal{I},\mathcal{Z}}\right)$$

- Also need measures for price impacts with and without the new data
  $\Rightarrow$ Calibrate constant price impact

Measurement

- Focus on 5 portfolios: Small, Large, Growth, Value, and S&P
- Data to value: IBES Forecasts for each of this five portfolios
- Estimation procedure is clear and easy to replicate
- Interesting patterns
- Can we do more? What drives the value of data?

Is more precise data more valuable?

	Investment Style				
	Small	Large	Growth	Value	All
Perfect Competition					
Investor with \$500,000 Wealth	0.00	\$1.7k	\$2.5k	\$490	\$3.5k
Investor with \$250m Wealth	0.00	\$566k	\$844k	\$164k	\$1.2m
With Price Impact					
Investor with \$500,000 Wealth	0.00	\$1.6k	\$2.5k	\$410	\$1.4k
Investor with \$250m Wealth	0.00	\$24k	\$57k	\$1.5k	\$253k
Informativeness of Forecasts	0.002	0.069	0.083	0.022	0.082

- Informativeness = precision of data as a signal of portfolio return
  - follow Dávila and Parlatore (2022) to estimate it

What drives the value of data?

- Data is valuable because it allows us to act on it
  - allows investors to adjust their portfolio from $a(\mathcal{I})$ to $a(\mathcal{I},\mathcal{Z})$
- For competitive investors restricted to one portfolio

$$\Delta \approx \frac{1}{2r\rho}\left[\underbrace{\left(\frac{\mathbb{E}[\Pi]^2}{\mathbb{V}[\Pi|\mathcal{I}]}+\frac{\mathbb{V}[\Pi]}{\mathbb{V}[\Pi|\mathcal{I}]}\right)}_{\approx\text{Profit}}\underbrace{\left[\frac{a(\mathcal{I},\mathcal{Z})-a(\mathcal{I})}{a(\mathcal{I})}\right]}_{\text{Portfolio adjustment}}\right]$$

- Decompose value of data into
  - Profit: independent of data to be valued!
  - Portfolio adjustment = how does the investor use the data
- What drives the heterogeneity in the value of data across investor types?

Value of data

	Investment Style				
	Small	Large	Growth	Value	All
Perfect Competition					
Investor with \$500,000 Wealth	0.00	\$1.7k	\$2.5k	\$490	\$3.5k
Investor with \$250m Wealth	0.00	\$566k	\$844k	\$164k	\$1.2m
Decomposition					
Profit	1.28	1.25	1.38	1.02	1.077
Change in risky holdings (%)	0.29	7.44	9.12	2.27	9.022

- Heterogeneity is coming from the use of data!

Summary

- Very interesting paper on very important question: what is the value of data?
- Methodology is clear and can be adapted to multiple sources of information

Summary

- Very interesting paper on very important question: what is the value of data?
- Methodology is clear and can be adapted to multiple sources of information
- How to interpret the estimates?
  - Lower bound for value of data for an investor with same priors as the econometrician's
  - Explore how tight the bound is (?)
- Interesting patterns across wealth levels, investor types, market power
- What drives these patterns?
  - Decompose value in interpretable objects
  - Counterfactuals, information choice (?), use additional data
- Robustness exercises: calibration of ρ and price impact are important!